

Modelado y visualización de relaciones entre contaminantes del aire y salud pública

Selene Berenice Prado Prado
Ing. en Tecnología de Software

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

7 de mayo de 2022

Índice

1 Introducción

- Hipótesis
- Objetivos

2 Antecedentes

- Monitoreo de calidad del aire
- Series de tiempo
- Clasificación de enfermedades
- Regresión lineal

- Regresión lineal múltiple

3 Estado del arte

4 Solución propuesta

- Herramientas
- Implementación de la solución

5 Experimentos

- Niveles de PM10
- Niveles de PM2.5

6 Conclusiones

Introducción

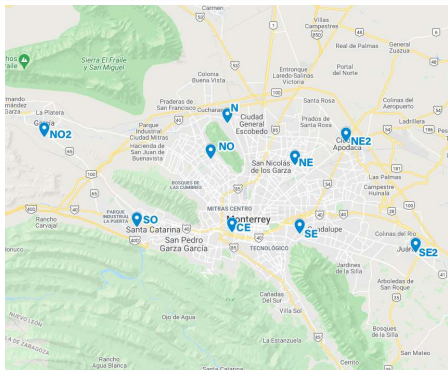


Figura 1: Estaciones de monitoreo de la calidad del aire pertenecientes al Sistema Integral de Monitoreo Ambiental (SIMA) [13].

Hipótesis

Hipótesis

Los modelos de regresión permiten obtener gráficos donde se pueden observar las relaciones entre el número de ingresos hospitalarios y los niveles de contaminantes del aire.

Objetivos

Objetivo general

Generar, implementar y evaluar modelos que muestran las relaciones existentes entre contaminantes del aire y salud pública tiene la finalidad de apoyar a la implementación de estrategias que aminoran los efectos negativos de los contaminantes del aire en la salud de las personas. Con los modelos generados se puede tener una herramienta que permite identificar gráficamente las relaciones con solo proporcionarle el conjunto de datos.

Monitoreo de calidad del aire

Contaminantes del aire

- Monóxido de carbono (CO)
- Dióxido de azufre (SO₂)
- Óxidos de nitrógeno (NO_x)
- Ozono (O₃)
- Partículas de tamaño menor a 10 micrómetros (PM10)
- Partículas de tamaño menor a 2.5 micrómetros (PM2.5)

Series de tiempo

Definición de serie de tiempo

Las series de tiempo se pueden definir como un conjunto de observaciones tomadas en un tiempo t determinado. Los estudios de series de tiempo relacionan estadísticamente los cambios temporales en la repercusión de cambios en la concentración de un contaminante en la población [3].

Serie de tiempo

Definición de semana epidemiológica

Una semana epidemiológica es un estándar de medición temporal que se utiliza para comparar datos en ventanas de tiempo definidas. La primera semana epidemiológica del año termina el primer sábado de enero de cada año [1].

Clasificación de enfermedades

Clasificación Internacional de Enfermedades (CIE)

Existe un instrumento estadístico y sanitario para identificar enfermedades llamado Clasificación Internacional de Enfermedades (CIE) que agrupa enfermedades en epidémicas, generales, locales ordenadas por origen geográfico, trastornos del desarrollo y lesiones [11].

Regresión lineal

Definición de regresión lineal

La tendencia w_0 de una serie de tiempo puede ser obtenida a partir de una regresión lineal de la misma [4]. Una regresión lineal es una metodología inferencial supervisada que busca predecir valores y dado un vector de variables de entrada t por medio del ajuste de coeficientes w de la función lineal:

$$\hat{y}(t, w) = w_0 + w_1x_1 + \dots + w_tx_t. \quad (1)$$

Regresión lineal múltiple

Definición de regresión lineal múltiple

Un modelo de regresión múltiple es un modelo complemento de la regresión lineal simple, el cual tiene dos o más variables independientes k que pueden influir en una variable dependiente y . Peniche-Camps y Cortez-Huerta [12] expresan la regresión múltiple mediante la siguiente ecuación:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon. \quad (2)$$

Estado del arte

Cuadro 1: Comparación de trabajos frente al desarrollado, donde ✓ indica que cumple con esta característica y × no cumple con esta característica.

Trabajo	Modelos de regresión lineal	Modelos de predicción	Evaluación de modelos	Estudio de contaminantes del aire	Estudio de problemas de salud
Martín y Bayle [10]	✓	×	×	✓	×
Guarnaccia <i>et al.</i> [5]	×	✓	✓	✓	×
Julia <i>et al.</i> [7]	✓	✓	×	✓	×
Zhang <i>et al.</i> [15]	✓	✓	×	✓	✓
Kim <i>et al.</i> [8]	✓	×	×	✓	✓
Liu <i>et al.</i> [9]	×	×	✓	✓	✓
To <i>et al.</i> [14]	×	×	×	✓	✓
Breton <i>et al.</i> [2]	✓	✓	×	✓	✓
Gupta <i>et al.</i> [6]	✓	✓	×	✓	✓
El presente trabajo	✓	✓	✓	✓	✓

Solución propuesta

Solución propuesta

La solución propuesta se compone de cuatro fases principales: **recolección de datos, selección y agrupación de datos, visualización de la evolución de las variables e implementación de modelos.**

Herramientas

Cuadro 2: Herramientas utilizadas.

Herramienta	Versión	URL
Python	3.8.8	https://www.python.org/
Jupyter Notebook	6.3.0	https://jupyter.org/
Imageio	2.9.0	https://imageio.readthedocs.io/
Latextable	0.2.1	https://pypi.org/project/latextable/
Matplotlib	3.3.4	https://matplotlib.org/
NumPy	1.20.1	http://www.numpy.org/
Pandas	1.2.4	https://pandas.pydata.org/
Seaborn	0.11.1	https://seaborn.pydata.org/
Scikit-learn	0.24.1	https://scikit-learn.org/
SciPy	1.6.2	https://docs.scipy.org/
Statsmodels	0.12.2	https://www.statsmodels.org/
Texttable	1.6.4	https://pypi.org/project/texttable/

Implementación de la solución

Visualizaciones generadas

Los tipos de visualizaciones generadas son:

- Series de tiempo.
- Gráficos de radar.

Modelos generados

Los tipos de modelos generados son:

- Modelos de regresión lineal.
- Modelos de regresión lineal múltiple.

Niveles de PM10

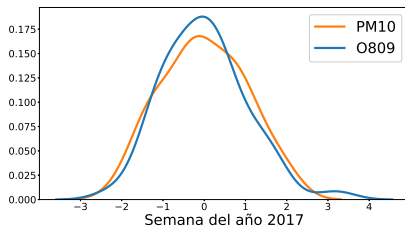


Figura 2: Evolución de los niveles de PM10 y el número de egresos con la CIE O809 en el 2017.

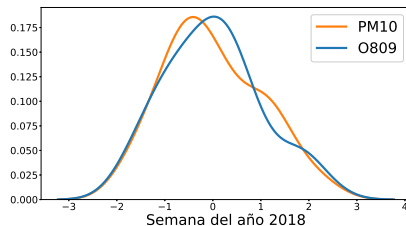


Figura 3: Evolución de los niveles de PM10 y el número de egresos con la CIE O809 en el 2018.

Cuadro 3: Resultados obtenidos PM10 2017

CIE	ρ	R^2	Valor p	ϵ
O809	-0.275	0.061	0.121	0.239
O829	0.100	0.091	0.055	0.287
O759	-0.085	0.116	0.029	0.294
O069	-0.247	0.070	0.094	0.222
K802	0.044	0.005	0.658	0.282

$\epsilon = \text{RMSE}$ para medir el error

PM10 2017

Cuadro 4: Resultados regresión lineal múltiple PM10 2017

Variable Dep.:	y	R²:	0.313			
Modelo:	OLS	Método:	Mínimos cuadrados			
Error:	0.226					
	coef	std err	t	P> t 	[0.025	0.975]
const	0.5859	0.146	4.002	0.000	0.289	0.883
O809	-0.3019	0.122	-2.472	0.018	-0.550	-0.054
O829	0.2685	0.123	2.185	0.036	0.019	0.518
O759	-0.2229	0.182	-1.222	0.230	-0.593	0.147
O069	-0.1441	0.120	-1.200	0.238	-0.388	0.100
K802	-0.0456	0.133	-0.344	0.733	-0.315	0.224

PM10 2018

Cuadro 5: Resultados obtenidos PM10 2018

CIE	ρ	R^2	Valor p	ϵ
O809	-0.271	0.015	0.440	0.275
O829	0.282	0.069	0.098	0.249
K802	0.277	0.084	0.066	0.152
O342	-0.401	0.100	0.044	0.248
N40X	-0.009	0.000	0.964	0.243

ρ = Coeficiente de correlación de Pearson

ϵ = RMSE para medir el error

PM10 2018

Cuadro 6: Resultados regresión lineal múltiple PM10 2018

Variable Dep.:	y	R²:	0.182			
Modelo:	OLS	Método:	Mínimos cuadrados			
Error:	0.159					
	coef	std err	t	P> t 	[0.025	0.975]
const	0.3885	0.210	1.849	0.073	-0.038	0.815
O809	-0.0114	0.188	-0.061	0.952	-0.392	0.370
O829	0.0854	0.214	0.400	0.692	-0.349	0.520
K802	0.3088	0.167	1.852	0.072	-0.030	0.647
O342	-0.1708	0.208	-0.820	0.418	-0.593	0.252
N40X	-0.1239	0.167	-0.740	0.464	-0.464	0.216

Niveles de PM2.5

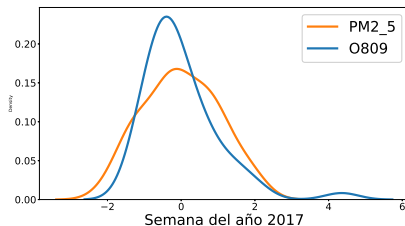


Figura 4: Evolución de los niveles de PM2.5 y el número de egresos con la CIE O809 en el 2017.

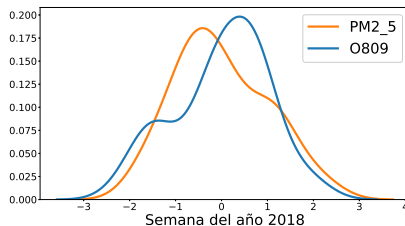


Figura 5: Evolución de los niveles de PM2.5 y el número de egresos con la CIE O809 en el 2018.

PM2.5 2017

Cuadro 8: Resultados regresión lineal múltiple PM2.5 2017

Variable Dep.:	y	R²:	0.210			
Modelo:	OLS	Método:	Mínimos cuadrados			
Error:	0.175					
	coef	std err	t	P> t 	[0.025	0.975]
const	0.5345	0.158	3.373	0.002	0.213	0.856
O809	-0.1754	0.132	-1.327	0.193	-0.444	0.093
O829	0.0701	0.133	0.527	0.602	-0.200	0.340
O759	-0.2585	0.197	-1.309	0.199	-0.659	0.142
O069	-0.2148	0.130	-1.652	0.108	-0.479	0.049
K802	-0.1164	0.144	-0.811	0.423	-0.408	0.175

PM2.5 2018

Cuadro 9: Resultados obtenidos PM2.5 2018

CIE	ρ	R^2	Valor p	ϵ
O809	-0.354	0.057	0.133	0.259
O829	0.225	0.046	0.177	0.256
K802	0.273	0.089	0.058	0.159
O342	-0.443	0.149	0.013	0.245
N40X	-0.074	0.001	0.842	0.241

ρ = Coeficiente de correlación de Pearson

ϵ = RMSE para medir el error

PM2.5 2018

Cuadro 10: Resultados regresión lineal múltiple PM2.5 2018

Variable Dep.:	y	R²:	0.259			
Modelo:	OLS	Método:	Mínimos cuadrados			
Error:	0.174					
	coef	std err	t	P> t 	[0.025	0.975]
const	0.7042	0.193	3.652	0.001	0.313	1.096
O809	-0.0863	0.172	-0.501	0.620	-0.436	0.264
O829	-0.1084	0.196	-0.553	0.584	-0.507	0.290
K802	0.3006	0.153	1.964	0.058	-0.010	0.611
O342	-0.3311	0.191	-1.733	0.092	-0.719	0.057
N40X	-0.2053	0.154	-1.336	0.190	-0.517	0.107

Conclusiones

Conclusión

En las series de tiempo se observa que la cantidad de egresos de la mayoría de las CIE estudiadas presentan una línea de evolución similar al contaminante PM10. Además, se observó que para el estudio de las relaciones entre los contaminantes y las CIE se obtuvieron mejores resultados con los modelos de regresión lineal múltiple frente a los modelos de regresión lineal simple.

Conclusiones

Trabajo a futuro

- Recolección de más datos de los niveles de contaminantes y de egresos para el estudio de años más recientes.
- Realizar un estudio de las relaciones de los niveles de determinados contaminantes y la cantidad de egresos por CIE empleando modelos de regresión lineal múltiple.
- Creación de un mapa interactivo donde se pueda observar por región los niveles de los contaminantes y la cantidad de egresos.

Referencias I

- [1] ARIAS, J. R. (2006), «What is an epidemiological week and why do we use them», *Skeeter*, **66**(1), pág. 7.
- [2] BRETON, R. M. C., J. C. BRETON, M. DE LA LUZ ESPINOSA FUENTES, J. KAHL, A. A. E. GUZMAN, R. G. MARTÍNEZ, C. GUARNACCIA, R. DEL CARMEN LARA SEVERINO, E. R. LARA y A. B. FRANCAVILLA (2021), «Short-Term Associations between Morbidity and Air Pollution in Metropolitan Area of Monterrey, Mexico», *Atmosphere*, **12**(10), pág. 1352. doi:10.3390/atmos12101352.
- [3] BROCKWELL, P. J., P. J. BROCKWELL, R. A. DAVIS y R. A. DAVIS (2002), *Introduction to time series and forecasting*, Springer. ISBN 9780387216577.
- [4] DARLINGTON, R. B. y A. F. HAYES (2016), *Regression analysis and linear models: Concepts, applications, and implementation*, Guilford Publications. ISBN 9781462521135.
- [5] GUARNACCIA, C., J. G. C. BRETON, R. M. C. BRETON, C. TEPEDINO, J. QUARTIERI y N. E. MASTORAKIS (2018), «ARIMA models application to air pollution data in Monterrey, Mexico», en *AIP Conference Proceedings*, tomo 1, AIP Publishing LLC, págs. 1–5.
- [6] GUPTA, A., H. BHERWANI, S. GAUTAM, S. ANJUM, K. MUSUGU, N. KUMAR, A. ANSHUL y R. KUMAR (2021), «Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models», *Environment, Development and Sustainability*, **23**(4), págs. 6408–6417. doi:10.1007/s10668-020-00878-9.
- [7] JULIA, A., D. A. FC LICHTENFELS, K. VAN DER PLAAT, C. C. DE JONG, D. S. VAN DIEMEN, I. POSTMA, C. M. NEDELJKOVIC, N. VAN DULJN, S. AMIN, M. LA BASTIDE-VAN GEMERT, D. VRIES *et al.* (2018), «Long-term air pollution exposure, genome-wide DNA methylation and lung function in the LifeLines cohort study», *Environmental health perspectives*, **126**(2), pág. 027 004. doi:10.1289/EHP2045.
- [8] KIM, J. S., Z. CHEN, T. L. ALDERETE, C. TOLEDO-CORRAL, F. LURMANN, K. BERHANE y F. D. GILLILAND (2019), «Associations of air pollution, obesity and cardiometabolic health in young adults: The Meta-AIR study», *Environment international*, **133**, págs. 105–180. doi:10.1016/j.envint.2019.105180.

Referencias II

- [9] LIU, Y., J. PAN, H. ZHANG, C. SHI, G. LI, Z. PENG, J. MA, Y. ZHOU y L. ZHANG (2019), «Short-term exposure to ambient air pollution and asthma mortality», *American journal of respiratory and critical care medicine*, **200**(1), págs. 24–32. doi:10.1164/rccm.201810-1823OC.
- [10] MARTÍN, R. M. y M. S. BAYLE (2018), «Impacto de la contaminación ambiental en las consultas pediátricas de Atención Primaria: estudio ecológico», en *Anales de Pediatría*, tomo 2, Elsevier, págs. 80–85.
- [11] ORGANIZATION, W. H. *et al.* (2016), «International Classification of Diseases. 2016», *World Health Organization*.
- [12] PENICHE-CAMPS, S. y M. CORTEZ-HUERTA (2020), «La costumbre al envenenamiento: El caso de los contaminantes atmosféricos de la ciudad de Guadalajara, México», *Revista de Ciencias Ambientales*, **54**(2), págs. 1–19. doi:10.15359/rca.54-2.1.
- [13] SIMA (2015), «Sistema Integral de Monitoreo Ambiental», <http://aire.nl.gob.mx>.
- [14] TO, T., J. ZHU, D. STIEB, N. GRAY, I. FONG, L. PINAULT, M. JERRETT, A. ROBICHAUD, R. MÉNARD, A. VAN DONKELAAR *et al.* (2020), «Early life exposure to air pollution and incidence of childhood asthma, allergic rhinitis and eczema», *European Respiratory Journal*, **55**(2). doi:10.1183/13993003.00913-2019.
- [15] ZHANG, Z., B. DONG, S. LI, G. CHEN, Z. YANG, Y. DONG, Z. WANG, J. MA y Y. GUO (2019), «Exposure to ambient particulate matter air pollution, blood pressure and hypertension in children and adolescents: a national cross-sectional study in China», *Environment international*, **128**, págs. 103–108. doi:10.1016/j.envint.2019.04.036.

¡Gracias por su atención!



Repositorio en línea